E/CONF.39/B/19
10 Oct.1962
ORIGINAL : ENGLISH

UNITED NATIONS CONFERENCE
ON THE APPLICATION OF SCIENCE AND TECHNOLOGY
FOR THE BENEFIT OF THE LESS DEVELOPED AREAS

Agenda Item : B.1

Sweden

NEW SURVEY TECHNIQUES APPLICABLE TO LABOR FORCE SURVEYS IN LESS DEVELOPED AREAS

By Tore DALENIUS,
Docent,
The Swedish Council for Social Science Research

Papers have been contributed to the United Nations Conference on the Application of Science and Technology for the Benefit of the Less Developed Areas as working papers for that Conference. They are published as presented by the contributors, and the contents and the views expressed are those of the contributors. (See notes overleaf.)

NOTES

1. Papers selected for inclusion in General Sessions will be available at the Conference in the following official languages of the United Nations: English, French, Russian and Spanish.

2. Papers selected for inclusion in Specialized Sessions will be available at the Conference in only one of the above official languages.

3. Summaries of all papers, as submitted by contributors, will be available at the Conference in the following official languages of the United Nations: English, French, Russian and Spanish.

4. Papers and summaries will not be available for general distribution in advance of the Conference. They will be published in advance in limited quantities only for the use of delegations of participating governments and of other authorized participants. Post-Conference publication and general distribution of the Conference Proceedings will follow United Nations practices as laid down by the United Nations Publications Board.

New survey techniques applicable to labor force surveys in less developed areas

by

Tore Dalenius, Docent

The Swedish Council for Social Science Research

1. Purpose and scope of the paper.

Considerable theoretical and methodological progress has taken place in the domain of statistical surveys since about 1950. The new techniques at the disposal of to-day's survey statisticians are largely of a world-wide applicability.

Among these new techniques there are, however, some which seem to be especially suitable for labor force surveys in less developed areas. It is the prime purpose of this paper to focus the attention of the reader on this particular aspect.

An exhaustive discussion of the subject is outside the compass of this short paper; the interested reader is advised to consult the references quoted in the text.

The aim of the paragraphs 2 and 3 is to provide the necessary institutional frame of reference for the presentation of four selected techniques in the following paragraphs.

2. Characteristic features of labor force surveys.

The term "labor force survey" will be used in this paper in a strictly technical sense. Thus, it will denote a system adapted in the United States, Canada and some other countries for the compilation of statistics of employment and unemployment by means of interview surveys. More specifically, it is characteristic of labor force surveys that

a. they are carried out as sample surveys;

b. they are repetitive surveys; they are carried out several times a year;

c. they make use of the panel technique, in combination with rotation sampling; and

d. the observations are collected by personal interviews.

For a general methodological review, reference[1] is given to International Labour Office (1948). Eckler (1961) should be consulted for a detailed description of a technically advanced application.

3. Crucial design conditions for statistical surveys in less developed areas.

It is significative of sample survey work in statistically advanced areas that it aims at the construction of practically (as distinguished from theoretically) optimum sample survey designs. For an illustration, the reader is referred to Hansen et al (1953); the discussion, p. 4-10, of "criteria for choice of good sample designs" is especially illuminating. The successful application of criteria such as those presented in this reference prerequisites the identification and evaluation of the "crucial design conditions" prevailing in the area under consideration.

We will assert here that sample survey work in less developed areas should, in principle, be based on the same criteria as those adopted in advanced areas.

In order to be able to illustrate the implication of this choice of criteria for sample survey work, we list here four points which are likely to appear on a list of crucial design conditions in less developed areas:

a. scarcity of well-trained statisticians, especially sample survey experts;

b. scarcity or lack of large-scale computers;

c. existing population registers are incomplete;

d. basic statistical information is unreliable.

---

1) Complete references are given in paragraph 9.

We will interpret this list to suggest that the choice of sample design should aim at

(i) the achievement of considerable technical and administrative simplicity; and

(ii) the application of techniques which permit an efficient use of inadequate resources.

In paragraphs 4-7, we will present four sample survey techniques of a recent origin which seem to be especially suitable for labor force sample surveys in less developed areas, where one or several of the four conditions listed above apply.

4. Replication.

The use of replication represents a feasible basis for sample survey work in areas with a scarcity of sample survey experts and lack of large-scale computers. For an extensive presentation of this technique, reference is given to Deming (1960). A hint of the basic idea will be given here by means of a simple example:

A sample survey is to be carried out to estimate the total number T of employed persons, 14 years of age or older. We select (to be specific) independently 10 "small" samples, replicates, and compute the 10 estimates

$$t_1, t_2, \ldots\ldots, t_5, \ldots\ldots, t_{10}$$

of T. Thereafter, we compute the overall estimate

$$\bar{t} = \frac{1}{10} \sum_{i=1}^{10} t_i$$

of T. The precision of t can be measured by the variance estimate:

$$s^2(\bar{t}) = \frac{1}{10} \cdot \frac{1}{9} \sum_{i=1}^{10} (t_i - \bar{t})^2$$

The computation of $s^2(\bar{t})$ is (usually) much simpler and faster than the corresponding computation, when replication is not used. It must be observed, however, that we must be prepared to pay a price for these gains in terms of decreased precision.

Replication may also be exploited for purposes of controlling non-sampling errors. References are given to Hansen et al (1961) and Lahiri (1958).

5. A technique for efficient use of an incomplete register.

In order to put probability sampling into practice, it is necessary to have access to a "frame", that is a list of the set of sampling units among which to select a sample; obviously the use of multi-stage sampling calls for a corresponding hierarchy of frames.

It is often, and rightly, emphasized that in order to be satisfactory, the frame has to be (virtually) complete. This should, however, not be interpreted to mean that an incomplete list of units cannot be effectively exploited in the sample selection procedure. We will briefly outline a suitable technique for such a use by means of a specific example.

At the final stage of a certain multi-stage sample design, a sample of dwelling units is to be selected from among the 14 dwelling units of a block. There exists a "dwelling unit directory" for this block, but it is not complete.

Consider a unique rule for numbering the dwelling units of the block, e.g. by reference to street and street-number. Such a numbering, if carried out, would create a complete sequence

$$U_1, U_2, \ldots U_{10}, \ldots, U_{13}, U_{14}$$

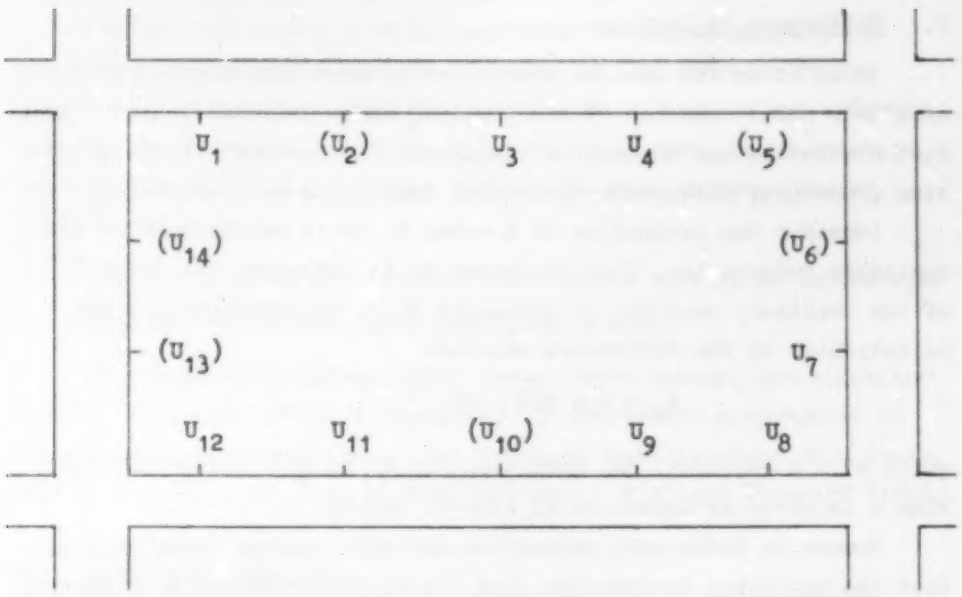
of dwelling units. However, instead of the complete sequence above, we have the following incomplete sequence

$$U_1, (U_2), \ldots, (U_{13}), (U_{14})$$

where the parenthesis ( ) symbolizes that the unit concerned is not listed in the directory. The situation is illustrated by means of the map on the next page. As shown there, the indexing 1, 2, ... 14 has been made in a systematic fashion round the block.

For purposes of sample selection, a dwelling unit not listed in the directory will be linked with one and only one dwelling unit listed in the directory. The rule to use in this process calls for the following linking:

(U_2) is linked with U_1 ;
(U_5) is linked with U_4 ;
(U_6) is linked with U_4 ;
(U_{10}) is linked with U_9 ;
(U_{13}) is linked with U_{12} ; and
(U_{14}) is linked with U_{12}



Thus, a dwelling unit not listed in the directory is linked with the nearest preceding unit.

The actual selection of a sample of dwelling units where to make interviews may be arranged as follows. A sample of dwelling units is first selected from the directory; this sample may e.g. be made up by U_3, U_9 and U_{12}. The interviewer is then instructed to make interviews not only in these three dwelling units but also in those dwelling units linked with these units. Thus, (U_{10}) is to be included in the survey, as are (U_{13}) and (U_{14}).

For a realistic illustration of an application of this technique to a large-scale sample survey, reference is given to Jabine et al (1961).

The alternative approach to the frame problem in the case discussed is to have a separate pre-listing operation precede the selection of the sample of dwelling units, in order to create a complete list of dwelling units from which to select the sample. Such an operation may be expensive.

6. Difference estimation.

It is often the case in sample survey work that statistics are available about each one of the sampling units included in the frame. Such statistics may be used as "auxiliary information" in the estimation procedure; difference estimation represents one such means.

Consider the estimation of a total T. If no use is made of the auxiliary information, T is estimated by t; likewise, the total Y of the auxiliary variable is estimated by y. Alternatively, T may be estimated by the difference estimate

$$t_\theta = t + \theta(Y - y)$$

where θ is a constant. The condition for t_θ to be a better estimate than t is given in Hansen et al (1953), chapter 11.

Common to difference estimation and some related techniques is that the auxiliary information need not be either accurate or current in order to be fit for use. In order to bring out the significance of this fact, we consider a labor force survey, the sample design of which embraces the use of parishes as sampling units. For each such unit in the frame, the approximate number of inhabitants as of some previous year is known, as recorded in a census; the sum of these numbers is the total Y of the auxiliary variable and is estimated by y. The fact that difference estimation makes no demands for up-to-date, accurate statistics and is relatively simple to apply, makes this technique especially attractive for use in less developed areas.

The attention of the reader is also drawn to composite estimation, as discussed in Hansen et al (1955).

7. The surprise-stratum technique.

Suppose we are carrying out a labor force survey, making use of "area sampling". The design calls for the selection of dwelling units within a sample of blocks in cities and segments in non-urban areas.

For such a design to be effective, it is important to identify in advance all areas with a large population, in order to be able to exercise a control of their impact on the estimates. Such a control may make use of stratified sampling or sampling with probabilities proportional to estimated size.

Clearly, the identification mentioned above prerequisites up-to-date, reliable statistics. In spite of our efforts, it may happen that we overlook some large areas. If such areas are selected, they are appositely referred to as "surprise units". In an one-time survey, such units may be dealt with in two ways, neither one of which is attractive:

a. they are kept in the sample; this results in a large variance; or

b. they are "rejected"; this introduces a bias.

In a repetitive survey like a labor force survey, the surprise-stratum technique offers a statistically attractive alternative to the two ways mentioned above. For technical details about this technique, reference if given to Kish and Hess (1959) and Woodruff (1959).

8. Some final remarks.

Because of space limitations, the previous discussion has concentrated on four selected new techniques which should find important applications in sample survey work in less developed areas. For a much fuller presentation of recent advances in sample survey theory and methods, reference is given to Dalenius (1962).

9. Bibliography.

Dalenius, T., Recent advances in sample survey theory and methods, Annals of Mathematical Statistics, 33, 325-349 (1962).

Deming, W.E., Sample Design in Business Research, Wiley, New York (1960).

Eckler, A.R., The continuous population and labour force survey in the United States. Chapter XIII in International Labour Office: Family Living Studies. A Symposium, Geneva (1961).

Hansen, M.H., Hurwitz, W.N. and Bershad, M., Measurement errors in censuses and surveys, Bulletin de l'Institut International de Statistique, 38, No. 2, 359-374 (1961).

Hansen, M.H., Hurwitz, W.N. and Madow, W.G., Sample Survey Methods and Theory, 1 and 2, Wiley, New York (1953).

Hansen, M.H., Hurwitz, W.N., Nisselson, H. and Steinberg, J., The redesign of the census current population survey, Journal of the American Statistical Association, 50, 701-719 (1955).

International Labour Office, Employment, Unemployment and Labour Force Statistics, A study of Methods, Geneva 1948.

Jabine, T.B., Hurley, R. and Hurwitz, W.N., Sample design and estimation procedure for the 1960 sample survey of agriculture, Revue de l'Institut International de Statistique, 29, No. 3, 1-12 (1961).

Kish, L. and Hess, I., Some sampling techniques for continuing survey operations, Proceedings of the Social Statistics Section American Statistical Association, 139-143 (1959).

Lahiri, D.B., Recent developments in the use of techniques for assessment of errors in nation-wide surveys in India, Bulletin de l l'Institut International de Statistique, 36, No. 2, 71-93 (1958).

Woodruff, R.S., The use of rotating samples in the Census Bureau's monthly surveys, Proceedings of the Social Statistics Section American Statistical Association, 130-138 (1959).

566 E - Oct.1962 - 1,600